LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-491036

Practical data mining and machine learning for optics applications: introduction to the feature issue

G. Abdulla, A. Awwal, K. Borne, T. Ho, W. Vestrand

July 20, 2011

Applied Optics

Disclaimer

Introduction to the Applied Optics feature issue

Ghaleb Abdulla, Abdul Awwal, Kirk Borne, Tin Kam Ho, and W. Thomas Vestrand

The world is becoming more instrumented every day, and data are being collected from embedded sensors to help understand our environment and to interact with it in better and more intelligent ways. The amount of data collected from these sensors is growing very fast, faster than the tools used to handle these data. Data contain information and encode knowledge about the phenomena being studied with sensors and instruments. The purpose of data collection is to use the data, and the corresponding acquired information and knowledge, to enhance a process, to predict the future, or to have a deeper understanding of a phenomenon, among other objectives. This special issue is an attempt to bridge the gap between the data mining and optics communities by introducing data mining techniques to the optics community.

Data mining (DM) is defined as the nontrivial discovery of novel, valid, comprehensible, and potentially useful patterns from data. The success of DM and machine learning (ML) techniques in exploring hidden patterns in large data sets triggered the utilization of such techniques in many fields including biology, astronomy, energy, economics, and text analysis. DM and ML algorithms have roots in statistics and mathematics. DM is focused on finding the right hypothesis. It utilizes search techniques to explore hidden patterns and correlations in the data, which otherwise require a tremendous amount of human time to explore.

The paper on "Autonomous subpixel satellite track end point determination for space-based images" by Lance M. Simms describes an algorithm to determine satellite track end points with subpixel resolution in spaced-based images. Image processing techniques are used to pre-process the data in order to clean the image. The algorithm accounts for satellite tracks that are curved by using an iterative and weighted least square fit to localized track end points.

The paper titled "Analysis of 1w Bulk Laser Damage in KDP" by David Cross and Christopher Carr examines the influence of laser parameters on laser induced damages in the Bulk of KDP crystals. Their analysis shows that the density of damage sites produced by 1053 nm light is less sensitive to pulse duration than was previously reported for 526 nm and 351 nm light. In addition to the analysis results, the paper describes a method for measuring the size and locations of thousands of KDP bulk sites and relating them to laser conditions.

The paper "Exploration of the multi-parameter space of ns-laser damage growth in fused silica optics" by Negres *et al.* explores the growth behavior of damage sites on the exit surface of fused silica optics. Techniques such as fitting data to statistical distributions are used to help understand and predict probability of growth. Clustering is used to explore the multi-parameter growth space and to identify outliers. Outlier sites (fast growers) identification will enhance the accuracy of predictive models that use distribution fitting of models based on average behavior.

The paper titled "Optical Modeling in Testbed Environment for Space Situational Awareness (TESSA)"by Sergie Nikolaev discusses building a library to simulate optical sensors. It describes a driver application for the simulation work which is the Testbed Environment for Space Situational Awareness a system that is used to analyze and answer Space Situational awareness questions. The author describes the

system where the developed libraries will be used and discusses the process to generate the simulated images using the understanding of the modeled optical sensors.

Although the set of papers is relatively small, the diversity of the topics discussed and the techniques used to understand the optics or the data generated by the optics is still apparent. This topic should be revisited in a few years since data are becoming the main product of all sensors. Understanding and making use of the data are becoming more important. We hope this issue will spur new activities and collaboration in this emerging field.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.